

基于可见性图网络的中国专利申请关注度分析



张梦月 胡军 严冠 李慧嘉

中央财经大学管理科学与工程学院 北京 102206

(zmy_go@163.com)

摘要 专利是创新的重要体现,很多人在进行专利申请之前会在网上对专利申请的过程进行查询,了解专利申请的步骤,这些人的搜索事实也是了解创新企业或个人对创新是否重视的一个手段。文中从一个全新的时间序列分析的视角即网络的角度,分析了关键字为“专利申请”的百度搜索指数时间序列的动力学特征。利用可见性图算法的原理将百度搜索指数时间序列转化为复杂网络,并计算其参数,分析其网络的拓扑结构。首先,通过计算2019年各省复杂网络的参数发现各省的专利关注度具有一定差异;其次,研究表明大多数网络均为无标度网络,原始时间序列具有分形的特征;最后通过聚类,可根据复杂网络的参数把31个省分为3类。文中分析了2011—2018年全国的百度搜索指数数据,通过社团结构的划分,可以发现时间序列的周期和中心节点对搜索指数影响的范围。

关键词 专利申请;关注度;百度指数;可见性图算法;复杂网络

中图分类号 TP393

Analysis of China's Patent Application Concern Based on Visibility Graph Network

ZHANG Meng-yue, HU Jun, YAN Guan and LI Hui-jia

School of Management and Engineering, Central University of Finance and Economics, Beijing 102206, China

Abstract Patent is an important embodiment of innovation. Many people will inquire about the process of patent application on the Internet and learn the application steps before patent application. In fact, the online searching is also a way to know whether innovative enterprises or individuals attach importance to innovation. This paper analyzes the dynamic characteristics of Baidu search index time series with the keywords of “patent application” from the perspective of a new time series analysis, namely from the perspective of network. The time series of Baidu search index is transformed into a complex network by using the principle of visibility graph algorithm, and its parameters are calculated to analyze the topological structure of the network. Firstly, by calculating the complex network, it can be found that the patent attention of each province has certain differences. Secondly, the study shows that most of the networks are scale-free networks and the original time series have fractal characteristics. Finally, by clustering, 31 provinces can be divided into 3 categories according to the characteristics of complex networks. This paper analyzes the data of Baidu search index from 2011 to 2018. By dividing the community structure, the time series period and the central node's influence on the search index can be found.

Keywords Patent application, Concern, Baidu index, Visibility graph algorithm, Complex network

1 引言

党的十八大明确提出“科技创新是提高社会生产力和综合国力的战略支撑,必须摆在国家发展全局的核心位置”。强调要坚持走中国特色自主创新道路、实施创新驱动发展战略。科技进步是国家和地区经济发展的重要推动力,科技创新和整个社会的经济状况有着密切的关系。我国一直把创新放在国家发展的首要位置。

知识产权能调动人们的创造主动性,促进社会资源的优化配置,是科技创新的重要支撑。而专利是知识产权的核心,专利的蓬勃发展有利于科技进步和经济发展,所以专利是一

个很好的技术进步或创新测度的指标,被国内外许多学者所推崇,其分析结果也被广泛应用在社会各界的各种评价与决策中。2016年,国家“十三五”规划纲要中明确将每万人口发明专利拥有量列为经济社会发展的主要指标之一,国务院印发的《“十三五”国家知识产权保护和运用规划》中再次提及提高专利质量和效益。

在中国,专利指国家专利主管机关授予的专利权、取得专利权的发明创造及国家颁发的授予专利权的专利证书。目前我国的专利可以分为3种类别——外观设计专利、发明专利和实用新型专利,其中,实用新型专利和外观设计专利的技术含量较低,而发明专利更具创新性,技术含量也最高。

收稿日期:2020-03-01 返修日期:2020-04-08 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(71871233);北京市自然科学基金(9182015)

This work was supported by the National Natural Science Foundation of China(71871233) and Beijing Natural Science Foundation(9182015).

通信作者:李慧嘉(Hjli@amss.ac.cn)

近年来,国家对专利越来越重视,开始采取专利资助、严打专利侵权等一系列措施,同时,社会也时常出现一些专利申请相关热点事件。但这些措施及热点对加强企业或个人对专利申请重视程度的效果不尽相同。因此,研究关键字为“专利申请”的百度搜索指数的波动特征及其影响机制,对帮助国家采取更有效的措施和政策,从而提高民众对创新的重视程度有重要的现实意义。

2 文献综述

近年来,随着国家对专利越来越重视,各省纷纷出台各种专利资助政策,引起了国内外许多学者的关注。Yuan 等基于政治关联与技术创新的视角发现企业政治关联阻碍了企业创新活动,降低了创新效率^[1]。Cheng 等通过实证证明了风险投资对我国技术创新有促进作用^[2]。Hou 等通过计量分析的方法研究外商直接投资对区域创新能力的影响^[3]。Chen 等发现各地区研究与试验发展经费内部支出及规模以上工业企业个数对发明专利数都有正面影响^[4]。Zhang 等以上海市专利资助政策为研究对象,用非参数统计方法分析了该政策对专利申请数量和质量的影 响^[5]。Yue 等研究了人均 GDP、企业产出比重和专利制度等因素和专利发展之间的关系^[6];有学者研究了 FDI、科技人力资源、科技经费支出、外资企业出口比重和人均 GDP 等因素对专利发展的影响^[7]。

Zhou 等通过建立国内专利申请受理情况时间序列的多种自回归模型和多种条件异方差模型,对专利受理量进行预测^[8]。Chen 等不再使用单一预测模型,而运用小波包多分辨率分解和重构技术,将专利申请量时间序列分解为同尺度的趋势项、季节项、循环项和不规则项等 4 项并进行单支重构,针对重构各项的统计特征分别采用 Autoregressive Integrated Moving Average 和 Autoregressive Moving Average 模型进行建模,得到专利申请量的预测值^[9]。

目前,大多数的专利研究主要有两方面:1)研究政治环境、投资环境和竞争环境等不同因素对专利发展的影响;2)使用不同的模型,利用专利申请受理量、授权量时间序列数据进行预测。很少有学者对专利申请的时间序列数据动力学特征进行分析。

本文使用关键字为“专利申请”的百度搜索指数,分析中国各省专利申请关注度的变化特点。

为了挖掘时间序列的更多信息,本文通过可见性图算法,把关键字为“专利申请”的百度搜索指数的时间序列转化为可见性的复杂网络。这种转化已有很多研究。Zhang 等通过伪周期时间序列构建复杂网络来研究时间序列,并研究了构造的网络与原始时间序列之间的关系^[10]。Xu 等把时间序列映射到最近的邻居网络^[11]。Marwan 等引入了递归网络的概念^[12]。Lucasa 等介绍了针对时间序列的可见性图算法^[13]。可见性图算法既保留了时间序列的特征,又使时间序列与复杂网络的特性相联系^[14],使用该算法,可以根据构建的复杂网络特性确定所研究系统是否具有确定性和随机性等^[15]。目前,基于可见性图算法的时间序列分析已被运用于不同的领域。Jiang 森运用可见性图算法对健康人与充血性心力衰竭病人的 R-R 间期序列,以及进行冥想训练的受试者训练前和训练中的 R-R 间期序列进行分析比较^[16]。可见性图算法

还用于分析中国 7 个碳试点市场价格时间序列的相似性和异质性^[17]。

本文从时间序列的角度出发,通过可见性图算法的原理将关键字为“专利申请”的百度搜索指数时间序列数据转化为可见性的复杂网络,通过度分布、社团结构等指标分析时间序列的动力学特征及其影响机制。

3 概念

3.1 复杂网络的基本概念

一个具体网络可抽象为一个由点集 $V(G)$ 和边集 $E(G)$ 组成的图 $G=(V, E)$ 。节点数记为 $N=|V|$,边数记为 $M=|E|$ 。 $E(G)$ 中每条边都有 $V(G)$ 中的一对点与之对应。如果任意点对 (i, j) 与 (j, i) 对应同一条边,则称该网络为无向网络,否则为有向网络。如果给每条边都赋予相应的权值,那么该网络就称为加权网络,否则称为未加权网络。

3.2 复杂网络参数

平均路径长度、聚类系数和度分布最常用于衡量复杂网络。

3.2.1 平均路径长度

网络中两个节点 i 和 j 之间的距离 d_{ij} 指连接这两个节点的最短路径上的边数。网络中任意两个节点之间的距离的最大值称为网络的直径,记为 D ,即:

$$D = \max_{i,j} d_{ij} \quad (1)$$

网络的平均路径长度 L 指任意两个节点之间的距离的平均值,其中 N 为网络的节点数。

$$L = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}}{\frac{1}{2}N(N-1)} \quad (2)$$

3.2.2 聚类系数

若网络中的一个节点 i 有 k_i 条边将它和其他节点相连,则这 k_i 个节点被称为节点 i 的邻居。在这 k_i 个节点之间最多可能有 $k_i(k_i-1)/2$ 条边。聚类系数 C_i 指这 k_i 个节点之间实际存在的边数 E_i 和最多可能的边数 $k_i(k_i-1)/2$ 之比,即:

$$C_i = \frac{E_i}{\frac{1}{2}k_i(k_i-1)} \quad (3)$$

整个网络的聚类系数 C 指所有节点 i 的聚类系数 C_i 的平均值。

3.2.3 度分布

节点 i 的度 k_i 定义为与该节点连接的其他节点的数目。我们用 $p(k)$ 评估复杂网络的度分布,其中 k 表示节点度或连接数。如果度分布符合幂率分布 $p(k) \sim k^{-\gamma}$,则该网络为无标度网络^[18]。

3.3 可见性图算法原理

本文采用 Lacasa 等^[19]提出的可见性图算法进行网络构建。将中国 31 个省区市(不含港澳台)的关键字为“专利申请”的百度搜索指数时间序列转换为复杂网络。首先,将子系统 $x(t)$ 的离散时间序列数据对应为网络的节点,根据可视准则构建网络连边:在时间序列 $(x(t))$ 中的任意两点数据 (t^a, x^a) 和 (t^c, x^c) 之间可视即可建立连边。即对于两点 (t^a, x^a) 和

(t^c, x^c) 之间任意一点 (t^b, x^b) ,当 $t^a < t^b < t^c$ 时,满足下述公式即可建立连边:

$$x^b < x^a + (x^c - x^a) \frac{t^b - t^a}{t^c - t^a} \quad (4)$$

如图1所示,直方条的高度代表每个时间点的数据值,如果两个直方条的顶端相互可视,则对应的两点在图中箭头后的网络相连。

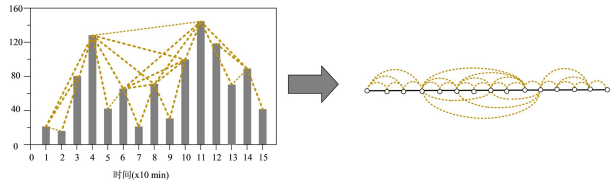


图1 可见性图算法

Fig.1 Visibility graph algorithm

其次,根据时间序列节点和边构建邻接矩阵,并形成网络图,如图2所示。

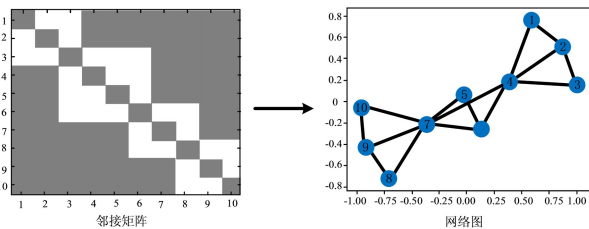


图2 可见性图网络的构建过程及网络特征提取

Fig.2 Visibility network construction process and network feature extraction

3.4 系统聚类

系统聚类法是先将聚类的样本或变量各自看成一群,然后确定类与类间的相似统计量,并选择最接近的两类或若干个类合并成一个新类,计算新类与其他各类间的相似性统计量,再选择最接近的两个群或若干群合并成一个新类,直到所有的样本或变量都合并成一类。常用的系统聚类法以距离为相似统计量来确定新类与其他各类之间的距离,如最短距离法及最长距离法、中间距离法、群平均法、离差平方和法、欧氏距离等^[20],本文采用欧氏距离进行聚类。两个 n 维向量 $\mathbf{a}(x_{11}, x_{12}, \dots, x_{1n}), \mathbf{b}(x_{21}, x_{22}, \dots, x_{2n})$ 间的欧氏距离为:

$$d_{12} = \sqrt{\sum_{k=1}^n (x_{1k} - x_{2k})^2} \quad (5)$$

3.5 社团划分算法

在复杂网络的研究当中,聚类结构的探测和分析已经成为一个非常重要的课题。一般来说,网络聚类指网络中的一组节点,相比网络的其他部分,其内部的相互关联更加紧密。如何从大规模的网络数据中发现最优的聚类结构是一个开放的热点问题,一般可以通过优化特定的指标函数来实现。模块度 Q 函数的形式如下:

$$Q = \sum_c \left[\frac{I}{m} - \left(\frac{2I + O_c}{2m} \right)^2 \right] \quad (6)$$

其中, m 表示网络中边的总数; I 表示两个端点均在同一社区中的边的数目; O_c 表示其中一个端点在社区中,而另一个端点在其他社区中的边的数目。 Q 的数值可以指示聚类结构的优劣, Q 值越大表示网络中的聚类结构越明显。

4 数据

本文数据来源于百度指数。百度指数(Baidu Index)是以百度海量网民行为数据为基础的数据分析平台,是当前互联网乃至整个数据时代最重要的统计分析平台之一,自发布之日便成为众多企业营销决策的重要依据。搜索指数是以网民在百度的搜索量为数据基础,以关键词为统计对象,科学分析并计算出各个关键词在百度网页搜索中搜索频次的加权和。根据使用百度搜索的不同来源,搜索指数被分为PC搜索指数和移动搜索指数。

随着互联网时代的到来,网络成为人们获取信息的重要途径,对于大多数人而言,在申请专利之前需要查阅大量资料来了解专利申请的步骤、所需要准备的资料和手续等,因此人们会通过网络查询专利申请的相关信息,从而百度搜索指数可以在一定程度上反映人们对专利的关注程度。

本文采用31个省区市(不含港澳台)的2019年1月1日—2019年12月31日、全国(不含港澳台)2011年1月1日—2018年12月31日关键词为“专利申请”的PC+移动端百度搜索指数分别进行分析。

5 各省数据的实证分析

5.1 中国各省可见性图网络的构建

本节采用可见性图算法原理对中国31个省区市(不含港澳台)的百度搜索指数数据进行网络构建,得到可见性图网络。该网络的各项指标如表1所列。

表1 各省可见性图网络参数

Table 1 Visibility network parameters

省份	平均度	直径	平均路径长度	密度	平均聚类系数
安徽	21.797	114	32.118	0.06	0.656
北京	14.679	73	14.995	0.04	0.637
福建	13.529	98	35.441	0.037	0.62
甘肃	13.299	150	49.621	0.037	0.489
广东	24.515	39	9.542	0.067	0.771
广西	13.863	134	42.748	0.038	0.567
贵州	10.981	125	38.02	0.03	0.548
海南	8.186	154	50.709	0.022	0.442
河北	15.233	76	27.719	0.042	0.698
河南	15.512	104	33.692	0.043	0.627
黑龙江	13.595	130	40.151	0.037	0.537
湖北	19.792	81	23.295	0.054	0.625
湖南	25.485	62	16.355	0.07	0.681
吉林	8.022	186	65.911	0.022	0.424
江苏	23.94	57	19.674	0.066	0.749
江西	11.288	119	33.908	0.031	0.623
辽宁	22.504	84	24.768	0.062	0.652
内蒙古	29.655	133	39.565	0.041	0.583
宁夏	10.959	133	41.822	0.03	0.525
青海	8.384	145	44.22	0.023	0.419
山东	23.063	66	26.196	0.063	0.745
山西	11.748	137	45.114	0.032	0.552
陕西	19.633	93	26.004	0.054	0.603
上海	35.178	73	13.688	0.097	0.639
四川	22.729	40	11.198	0.062	0.694
天津	12.197	112	31.22	0.034	0.558
西藏	6.274	198	67.178	0.017	0.283
新疆	9.984	145	45.944	0.027	0.472
云南	17.995	122	34.533	0.049	0.592
浙江	25.995	33	7.442	0.071	0.752
重庆	24.356	63	11.179	0.067	0.61

5.2 无标度网络:分形时间序列

图3展示了安徽省的复杂网络的幂率分布图。可以看出,安徽省的可见性图网络是无标度网络,连边较少的时间节点比较多,度值大的节点所占比例较小,随着度的增加,节点的数量减少。这反映出了原始时间序列的波动性。

度分布表明,安徽省专利申请的百度搜索指数是一个具有长期相关性的分形时间序列。一方面,原始时间序列在不同的时间范围内,如日、周、月等在统计数据上具有相似性;另一方面,由于长期相关性,未来安徽省的百度搜索指数变化可能会与过去的某个时期相似。

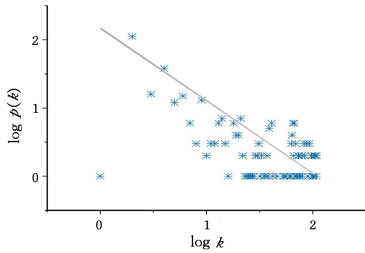


图3 安徽省可见性图网络的幂率分布

Fig. 3 Power distribution of visibility network in Anhui province

表2列出了中国31个省区市(不含港澳台)复杂网络的度分布情况,31个省区市(不含港澳台)的可见性图网络的度分布都为幂率分布。由表3的具体拟合结果可知,江西省和浙江省的 R^2 在0.75左右,其中,浙江省的 R^2 最低,其他省的 R^2 都高于0.8。其中,广东、江苏、山东、上海、浙江的幂指数小于1,其他的幂指数都在1~2之间,说明这5个省的原始时间序列数据波动性较小。除了这5个省,其他的可见性图网络都是无标度网络,对应的原始时间序列数据波动性大,且是具有长期相关性的分形时间序列。

表2 可见性图网络的幂率分布和拟合优度

Table 2 Power law distribution and goodness of fit of visibility networks

省份	Y	R	省份	Y	R
安徽	1.062	0.804	辽宁	1.017	0.837
北京	1.188	0.859	内蒙古	1.204	0.838
福建	1.225	0.847	宁夏	1.297	0.825
甘肃	1.362	0.859	青海	1.612	0.858
广东	0.972	0.820	山东	0.985	0.873
广西	1.252	0.871	山西	1.305	0.878
贵州	1.332	0.887	陕西	1.093	0.827
海南	1.559	0.865	上海	0.973	0.846
河北	1.114	0.846	四川	0.974	0.819
河南	1.175	0.865	天津	1.292	0.872
黑龙江	1.316	0.881	西藏	1.496	0.802
湖北	1.105	0.843	新疆	1.444	0.874
湖南	1.011	0.861	云南	1.188	0.851
吉林	1.596	0.8617	浙江	0.875	0.764
江苏	0.999	0.801	重庆	1.057	0.829
江西	1.113	0.786			

5.3 区域系统聚类

本文以可见性图网络的参数为聚类依据,采用系统聚类的方法对31个省区市(不含港澳台)的可见性图网络进行聚类。图4展示了聚类结果,每个节点表示成对聚类,两个节点之间的纵坐标长度表示两个合并簇之间的距离,每一类下的

省区市在2019年的专利申请关注度变化相似。

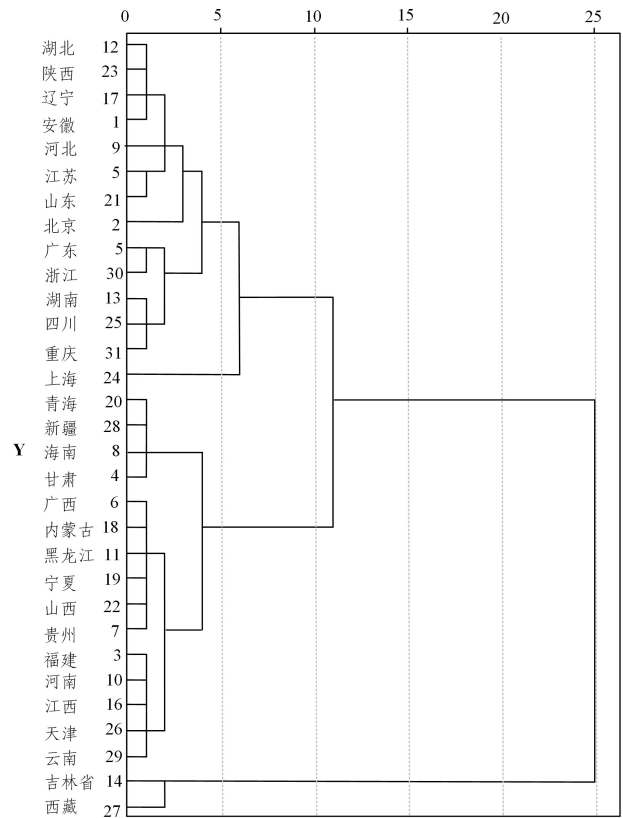


图4 系统聚类结果

Fig. 4 System clustering results

由图4可知,把31个省区市(不含港澳台)分为3类最佳,吉林和西藏被分为一组,称为第一组;青海、新疆、海南、甘肃、广西、内蒙古、黑龙江、宁夏、山西、贵州、福建、河南、江西、天津及云南被分为一组,称为第二组;湖北、山西、辽宁、安徽、河北、江苏、山东、北京、广东、浙江、湖南、四川、重庆及上海被分为一组,称为第三组。

专利申请的百度搜索指数能从一定程度上反映该省区市对专利申请的关注度变化特点。第一组的西藏、吉林网络的直径较大、边较少,平均度、聚类系数及密度都较低,表明节点之间的关系较弱,且这一类省区市的专利发展也相对落后。第二组的省区市的直径、边、平均度、聚类系数及密度都处于中间水平,节点之间具有一定联系,2019年专利申请的关注度较高。第三组的省区市包括北京、上海、广东、江苏、浙江等的网络边、平均度、聚类系数和密度等都较高,表明节点之间关系密切,2019年这一类省区市的专利申请关注度非常活跃。

6 全国数据的实证分析

6.1 专利申请关注度的可见性图网络构建

为了全面分析专利申请关注度时间序列的特性,本节采用可见性图算法原理分别对2011-2018年全局和局部的“专利申请”百度搜索指数时间序列构建可见性图网络。

我们根据“专利申请”百度搜索指数的历史数据划分了3个4年的时间序列窗口,每个窗口包含1461个节点,来构建局部的可见性图网络,并依次向后移动2年产生新的窗口,如图5所示。

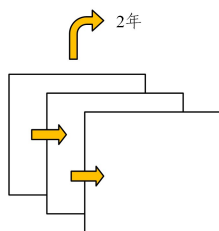


图5 时间窗口的移动原理

Fig. 5 Time window moving schematic

时间窗口大小的不同导致网络节点数不同,影响对应可见性图网络的性质。对于专利申请关注度时间序列来说,一个4年的时间窗口有较长的时间区间来构建合适规模的可见性图网络。

构建的全局和局部的可见性图网络的各项指标如表3所列。

表3 全国(不含港澳台)可见性图网络的参数

Table 3 National(excluding Hong Kong, Macao and Taiwan regions) visibility network parameters

	全局	2011—2014年	2013—2016年	2015—2018年
平均度	165.092	26.704	108.673	62.167
直径	112	104	37	89
平均路径长度	15.844	25.738	5.348	19.248
密度	0.057	0.018	0.074	0.043
平均聚类系数	0.745	0.75	0.821	0.744

6.2 社团结构

在每个时间窗口中,本文使用 Girvan-Newman 算法划分可见性图网络中的社团结构。社团检测的结果如表4和图6—图8所示,图中不同的颜色表示不同节点所属的社团。

从表4可知,第一个窗口和第三个窗口中的社团数目较接近,这说明在这两个时间段中对专利申请关注度的影响因素是相似的。第二个窗口中只有8个社团,说明在第二个时间窗口出现了较强的影响因素,导致其他因素的影响不大,使得更多的数据处于一个社团中。

表4 社团划分

Table 4 Community division

时间窗口	2011—2014年	2013—2016年	2015—2018年
社团数量	16	8	15

从图6—图8可知,专利申请相关的某些重大事件对应的节点所处的社团规模通常较大。中心节点对后续百度搜索指数影响的范围较广,但是从不同颜色的社团中也可以发现,不同中心节点影响持续的时间不同。

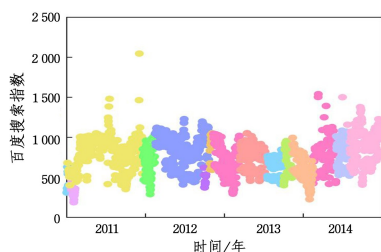


图6 2011—2014年可见性图网络的社团划分

Fig. 6 Community division of visibility networks in 2011—2014

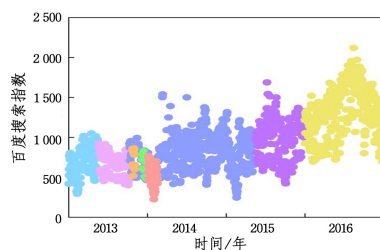


图7 2013—2016年可见性图网络的社团划分

Fig. 7 Community division of visibility networks in 2013—2016

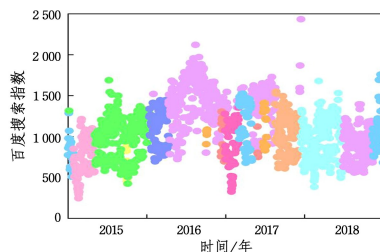


图8 2015—2018年可见性图网络的社团划分

Fig. 8 Community division of visibility networks in 2015—2018

同时,本文发现社团划分后,每个社团对应的时间序列周期并不是完全相连的,大社团中夹杂着小社团。例如,图8中2016年8月13日这个节点连接的社团并不是连续的时间序列,有些时间节点相隔较远。这说明专利申请相关的某些重大事件不仅会影响它附近的节点,还可能对相隔较远的时间序列数据产生影响。

结束语 本文使用31个省区市(不含港澳台)2019年1月1日—2019年12月31日、全国(不含港澳台)2011年1月1日—2018年12月31日“专利申请”的百度搜索指数,将原始时间序列通过可见性图算法转化为可见性图网络,研究其动力学特征,为分析专利申请的时间序列特征提供了全新的角度。

在对数据的实证分析中发现,中国2019年各省区市(不含港澳台)的专利申请关注度存在一定差异,其中,除了广东、江苏、山东、上海、浙江,其他都是无标度网络,原始百度搜索指数的时间序列是具有长期相关性的分形时间序列;最后,基于可见性图网络的参数进行聚类,进而把中国31个省(不含港澳台)分为3类。

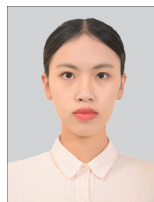
为了更好地分析百度搜索指数的波动,本文将时间序列划分为3个4年的窗口,分别在每个窗口用可见性图算法进行分析。通过社团划分发现,时间序列周期可以转化为可见性图网络中的社团,中心节点通常与大社团连接,说明中心节点对百度搜索指数波动的影响范围较广。

本文利用可见性图算法构建的网络是一个无向无权的网络,在转化过程中丢失了原时间序列的一些信息,导致其并不能很好地分析时间序列的变化特征,因此在未来可选取其他分析时间序列变化特征的方法进行优化。

参考文献

- [1] YUAN J G, HOU Q S, CHENG C. Curse Effect of Enterprise Political Resources—Based on the Study of Political Relevance

- and Enterprises Technological Innovation [J]. *Management World*, 2015(1):139-155.
- [2] CHENG K, LIU R H, LIU Y. Research on the Role of Venture Capital in China's Technological Innovation [J]. *Inquiry Into Economic Issues*, 2006(10):17-22.
- [3] HOU R X, GUAN J C. The Influence of Foreign Direct Investment on China's Regional Innovation Ability [J]. *China Soft Science*, 2006(5):104-111.
- [4] CHEN G H, LAN B J. Research on R&D Expenditure, Competition Degree and Regional Innovation Ability of China-An Empirical Analysis Based of Domestic Patent Applications and R&D Data from 1998 to 2004 [J]. *Economists*, 2007(3):101-106.
- [5] ZHANG Q H, LUO J W. Analysis on the Effect of Shanghai Patent Subsidy Policy on the Number of Patent Applications [J]. *Studies in Science of Science*, 2009, 27(5):682-685.
- [6] YUE H L. Patent laws and innovation in China[J]. *International Review of Law & Economics*, 2009, 29(4):304-313.
- [7] WU J, HU D, WANG Q. Role of education for spillover effects of FDI on innovation in China; Evidence from the 1998-2008 provincial panel data[C]// 2010 IEEE 17Th International Conference on Industrial Engineering and Engineering Management. IEEE, 2010:454-457.
- [8] ZHOU R F, YU J L. GARCH Model and Its Prediction of Time Series of Domestic Patent Application Acceptance [J]. *Journal of Zhongyuan University of Technology*, 2008(3):31-35.
- [9] CHEN Z Y, YANG D Z, MA W, et al. Prediction of Patent Application Quantity Based on the Combination Model of Wavelet Packet Analysis [J]. *Public Communication of Science & Technology*, 2011(17):226-227.
- [10] ZHANG J, SUN J F, LUO X D, et al. Characterizing pseudoperiodic time series through the complex network approach[J]. *Physica D: Nonlinear Phenomena*, 2008, 237(22):2856-2865.
- [11] XU X K, ZHANG J, SMALL M. Superfamily phenomena and motifs of networks induced from time series[C]// Proceedings of the National Academy of Sciences. 2008:19601-19605.
- [12] MARWAN N, DONGES J F, ZOU Y, et al. Complex network approach for recurrence analysis of time series [J]. *Phys. Lett. A*, 2009, 373(46):4246-4254.
- [13] DONNER R V, ZOU Y, DONGES J F, et al. Recurrence networks: A novel paradigm for nonlinear time series analysis[J]. arXiv:0908.3447.
- [14] LACASA L, LUQUE B, BALLESTEROS F, et al. From time series to complex networks: The visibility graph[C]// Proc. Natl. Acad. Sci. USA. 2008:4972-4975.
- [15] LI X, YANG D, LIU X, et al. Bridging time series dynamics and complex network theory with application to electrocardiogram analysis[J]. *IEEE Circuits Syst. Mag.*, 2012, 12(4):33-46.
- [16] JIANG S. Signal Analysis of Heart Rate Variation Based on Visibility Network Method [D]. Nanjing: Nanjing University, 2014.
- [17] GAO Z K, CAI Q, YANG Y X, et al. Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series [J]. *Sci. Rep.* doi:10.1038/srep35622.
- [18] WATTS D, STROGATZ S. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393:440-442.
- [19] LACASA L, LUQUE B, BALLESTEROS F, et al. From time series to complex networks: The visibility graph[C]// Proceedings of the National Academy of Sciences of the United States of America. 2008:4972-4975.
- [20] XIE B C, ZHU J P, WANG X Y. The Application of Excel in big Data Mining[M]. Xiamen University Press, 2016:104.



ZHANG Meng-yue, born in 1999, undergraduate. Her main research interests include data mining and operational research.



LI Hui-jia, born in 1985, distinguished research fellow. His main research interests include data mining, pattern recognition, complex networks, and control theory.